U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-258 847

The EPA Program for the Standardization of Stationary Source Emission Test Methodology - A Review

Environmental Monitoring & Support Lab, Research Triangle Pk, NC

Aug 76

PB 258 847

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

Research Triangle Park, North Carolina 27711

TECHNICAL REPORT DATA
Please read Instructions on the reverse before completing

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/4-76-044		

4. TITLE AND SUBTITLE	5. REPORT DATE
THE EPA PROGRAM FOR THE STANDARDIZATION OF STATIONARY SOURCE EMISSION TEST METHODOLOGY - A REVIEW	August 1976
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
M. Rodney Midgett	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
Same last 12	1HD621
	11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Environmental Monitoring and Support Laboratory Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, North Carolina 27711	Final - In-house
	14. SPONSORING AGENCY CODE
	EPA-ORD

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This report contains the results from a program designed to standardize those emission test methods promulgated by the EPA for use in determining compliance with Federal emission standards. The approach taken has been to conduct at least a limited laboratory and field evaluation, followed by an interlaboratory collaborative test of each method. Emphasis here is placed on the collaborative testing, the results of which are presented in terms of within-laboratory, between-laboratory, and laboratory bias standard deviations. These estimates are based on single-run results, and not on the results of three consecutive runs as would be required in conducting compliance testing. A brief discussion is given of the manner in which the precision estimates are derived. Determination of method accuracy is also considered where practical. The design of each test, deficiences in test designs, and other problems affecting the test results are discussed. An improved test design that overcomes most of the problems observed in earlier tests is described. A brief discussion of current projects and future plans is given as well as references to the numerous reports on the results of the methods standardization activities.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Air Pollution Sampling Evaluation	Collaborative Testing Methods Standardization (or Methods Evaluation) Stationary Sources Emissions Testing Methods	13B

PRICES SUBJECT TO CHANGE

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
Release to Public	Unclassified	54
	20. SECURITY CLASS (This page)	22. PRICE
	Unclassified	$4.50

EPA Form 2220-1 (9-73)

EPA-600/4-76-044
August 1976

THE EPA PROGRAM FOR THE STANDARDIZATION OF STATIONARY SOURCE EMISSION TEST METHODOLOGY - A REVIEW

by

M. Rodney Midgett
Quality Assurance Branch
Environmental Monitoring and Support Laboratory
Research Triangle Park, North Carolina 27711

ENVIRONMENTAL MONITORING AND SUPPORT LABORATORY
QUALITY ASSURANCE BRANCH
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27711

DISCLAIMER

This report has been reviewed by the Environmental Monitoring and Support Laboratory, U.S. Environmental Protection Agency, and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

CONTENTS

Section		Page
LIST OF TABLES		iv
ACKNOWLEDGMENTS		v
1.	INTRODUCTION	1
2.	CONCLUSIONS AND FUTURE PLANS	3
3.	THE METHODS STANDARDIZATION PROCESS	6
	Steps in the Standardization Process	6
	Design of Collaborative Tests	8
	Analysis of Collaborative Test Data	13
4.	RESULTS OF THE METHODS STANDARDIZATION PROGRAM	17
	Stack Gas Velocity and Volumetric Flow Rate	17
	Gas Analysis for Carbon Dioxide, Excess Air, and Dry Molecular Weight	18
	Moisture Fraction	21
	Particulates	22
	Sulfur Dioxide	27
	Nitrogen Dioxide	29
	Sulfuric Acid Mist/Sulfur Dioxide	30
	Opacity of Stack Emissions	32
	Carbon Monoxide	35
	Beryllium	36
5.	REFERENCES	43
TECHNICAL REPORT DATA AND ABSTRACT		46

LIST OF TABLES

Number		Page
1.	Methods Collaboratively Tested under the Methods Standardization Program	39
2.	Precision Estimates for Those Parameters Where Standard Deviation Was Proportional to the Mean Value, δ	40
3.	Precision Estimates for Those Parameters Where Standard Deviation Was Independent of the Mean Value, δ	41
4.	Collaborative Test of EPA Method 9	42

ACKNOWLEDGMENTS

The author wishes to acknowledge with appreciation the assistance of Dr. Henry F. Hamil, Southwest Research Institute, San Antonio, Texas, and Mr. Paul C. Constant, Jr., of Midwest Research Institute, Kansas City, Missouri, who had responsibility for planning and coordinating the testing under contract to the Environmental Protection Agency. Appreciation is also extended to the many other individuals and organizations who took part in the planning and execution of the tests, including those who participated as collaborators. A very special acknowledgment is due those organizations and company personnel who voluntarily made their plant facilities available as test sites and who otherwise cooperated in making these studies possible. These organizations and individuals are too numerous to list here, but each has been given proper acknowledgment in the individual collaborative test reports.

SECTION 1

INTRODUCTION

Under the authority of Section 111 of the Clean Air Act, as amended, the U.S. Environmental Protection Agency (EPA), on December 23, 1971, promulgated its first group of new source performance standards, which placed restrictions on the allowable emissions from new plants in five industrial categories.[1] These were followed by standards for seven additional industries,[2] and standards covering several others have now either been promulgated or are in varying stages of development. In addition, four substances (asbestos, beryllium, mercury, and vinyl chloride) have been designated hazardous air pollutants, and emission standards for the first three have been promulgated under Section 112 of the Clean Air Act.[3]

Of fundamental importance to enforcement of the above standards is the measurement process. The new source performance standards are set for a particular facility after first determining in existing well- controlled sources the emission limitation levels attainable using best available control technology, with consideration being also given to cost. A specific source test method is used to determine these emission levels, and that method in turn becomes the referece method for demonstrating compliance with the new source performance standard. At the time that the initial new source performance standards were established, many of these methods had not been fully evaluated, nor had their precision, accuracy, and general reliability in the hands of typical users been determined. It is for this reason that the Quality Assurance Branch, Environmental Monitoring and Support Laboratory, EPA, has for the past 3 years been engaged in a systematic program to standardize

or validate those source test methods that will be used to determine compliance with Federal emission standards.

Traditionally, the within-laboratory and between-laboratory precision of test methods is determined through collaborative testing (round-robin testing). The collaborative test is designed so that each participant makes one or more measurements on identical samples using the same test method. Then, from a statistical analysis of the results, an estimate is made of the within-laboratory and between-laboratory precision of the test method. This general technique has been used very widely for the validation of methods for the analysis of such items as water, drugs, food and agricultural products, fertilizers, coal, and ores.

Our experience has shown that before a stationary source test method can be successfully collaboratively tested, it must be described in sufficient detail to ensure that each collaborator uses exactly the same sampling and analysis procedures, and further, it must give repeatable results when one laboratory analyzes the same sample several times. This can only be determined through intensive method evaluation, which now constitutes a large portion of the total program. This includes a rigorous evaluation of the field sampling as well as the analytical aspects of the method prior to collaborative testing.

This standardization program has resulted in more fully described methods of known precision and accuracy and of proven reliability. With such reliable methodology, the agency should be in a better position to enforce compliance with the standards that are either now or soon to be in the Federal regulations.

SECTION 2

CONCLUSIONS AND FUTURE PLANS

Results obtained thus far from the methods standardization program indicate that the program has been successful for the most part, although questions still remain concerning the performance of some of the methods tested. We have shown that the methods for stack gas velocity and volumetric flow rate, particulates, sulfur dioxide, nitrogen oxides, and plume opacity (Methods 2, 5, 6, 7, 9-Ref. 1, pp. 24884-24885, 24888-24893, 24895) are indeed reliable if used properly under the conditions for which they were designed. The Orsat procedure (Method 3 - Ref. 1, pp. 24886-24887) is generally satisfactory provided its limitations are recognized and its limits of precision can be accepted. The method for carbon monoxide (Method 10 - Ref. 2, pp. 9319-9321) is thought to be capable of good accuracy and precision, but it appears that the suppliers of standard gases need to improve the state of their technology. The collaborative test also indicates that some users of nondispersive infrared (NDIR) instrumentation need further training in correcting for the nonlinear response characteristics of their instruments.

Results of the tests of the sulfuric acid mist/sulfur dioxide method (Method 8 - Ref. 1, pp. 24893-24895) indicate that either this method suffers from extremely poor precision, or the test design was incapable of compensating for the normal range of concentration and velocity variation in time and space at the selected test site. The cause of this poor precision has still not been found, but another test of the method, similar to the paired train test of Method 5 is planned as a future project.

Much of the imprecision and lack of accuracy observed in the test of the beryllium method (Method 104 - Ref. 3, pp. 8846-8850) seemed to occur in the analysis phase. But this was probably related more to collaborator

competency than to deficiences in the method itself. Atomic absorption is known to be a reliable technique for the analysis of beryllium in the absence of interferences or when such interferences can be eliminated. There are presently no plans for a future test of Method 104. But any such future test would likely include a more elaborate test design as mentioned above for Method 5.

Before leaving the subject of these methods, it should be pointed out that EPA Methods 1 through 8 have recently been revised and many of these revisions reflect refinements and improvements brought about through the methods standardization program. Other changes are the result of other experience gained within EPA since the initial promulgation in December 1971, as well as the present agency policy toward use of metric units. While the basic chemistry and procedures of these methods remain unchanged, the revisions supply much needed detail and correct other deficiencies of the original 1971 versions. They are due for proposal in a forthcoming issue of the Federal Register.

Several other test methods are now in various stages of the standardization process. The EPA method for determining mercury emissions from chlor-alkali plants (Method 102 - Ref. 3, pp. 8840-8845) has been evaluated and the analytical phase of the method has been collaborative tested.[4] The results indicate that many analysts have difficulty achieving satisfactory precision with the method, so a modified procedure has been developed, which eliminates many of the problems of the original Method 102. A collaboratove test of this modified procedure has shown conclusively that it is superior to the original. The agency is now planning to adopt this modified procedure as the official compliance test method.

The EPA method for determining the hydrogen sulfide content of petroleum refinery process fuel gases (Method 11 - Ref. 2, pp. 9321-9323) has been evaluated in the laboratory and found to suffer a major interference from thiols, which are

common constituents of such gas streams. A modified method, which is designed to eliminate this interference problem, was therefore developed, and this method is now being evaluated.[5] A full-scale collaborative test is planned for the near future; if successful, the method will be recommended as a replacement for the current Method 11.

The methods for fluorides (Methods 13A, 13B)[6] and vinyl chloride (Method 106)[7] are currently undergoing evaluation, although this work is still in its early stages. Field investigations of the fluoride methodology have indicated that the field sampling phase of Methods 13A and 13B is generally reliable.[8] These methods will also be submitted to inter-laboratory collaborative testing upon completion of the laboratory and field evaluations, provided that these evaluations prove them to be technically sound. Other methods will be introduced into the program as priorities dictate, and as time and funds permit.

SECTION 3

THE METHODS STANDARDIZATION PROCESS

STEPS IN THE STANDARDIZATION PROCESS

The validation of source test methodology is a complex, lengthy, and costly process, but years of experience have indicated the need for a complete and systematic examination even for those methods and measurement principles with fairly extensive histories of usage. Basically, this examination consists of the following steps:

First, the method is examined for technical accuracy, clarity, completeness of detail, etc. Regardless of how good the inherent capabilities of a measurement principle, a method may not give reliable results if it is poorly written, has errors in critical spots, or has such a scarcity of procedural detail that operations cannot be duplicated from one user to another. If the method is found to be deficient here, it may need to be rewritten.

Second, the method is subjected to a thorough and rigorous laboratory evaluation. This evaluation may include investigations of sample collection efficiency, applicable concentration range, mode of calibration, effects of interferences, etc. It may be said that these are the job of the researcher and should be done at the time the method is developed. This is true in principle, but experience has shown that many times such investigations were not conducted, or were carried out in such a superficial manner as not to uncover significant method deficiencies. This laboratory evaluation generally concludes with an experimentally designed ruggedness test[9] to determine the critical operational parameters of the method. The results of this evaluation may indicate the need to modify or rewrite the method.

Third, the method receives field evaluation at an applicable test site to determine its overall suitability for making the intended measurement. This evaluation may approach routine source testing, and is designed to check out the performance of the method under typical field conditions and to evaluate the source itself as a possible site for a collaborative test. When laboratory investigations by themselves are insufficient for determining the performance of certain source test methods, extensive field evaluations and statistically designed experiments using novel and original evaluation techniques may be required.[8,10] As before, results of this work may indicate the need for method modification or revision.

As the culmination to this chain of events, the method is submitted to an interlaboratory collaborative test at an appropriate test site using qualified participants to determine its precision, accuracy, and field reliability. Based on the test results and other information gained from the test, a final draft of the method is prepared and recommendation is made for its adoption by the agency. A report is also prepared documenting the test itself. Since collaborative testing is the major milestone event in validating the performance of a method as used by different individuals, it will be discussed separately in a subsequent section of this report.

The above sequence of steps represents a validation process that has evolved with time from a more simple and naive approach at the beginning of the program. When the standardization process was first begun for those methods that had already been published in the Federal regulations, the assumption was made that these were well written, fully described, and adequately researched methods needing no further evaluation. Therefore, steps 1 and 2 of this sequence were all but eliminated, and plans were made for method collaborative testing after limited field evaluation. This is

now considered to have been a mistake. While it is true that most of these methods are based on sound measurement principles that had been widely used prior to promulgation, it was found that procedure details were occasionally too sketchy to ensure that different users would execute procedural operations in a sufficiently similar manner, too many options were frequently allowed in the way certain operations were executed, and in a few instances, the method had faulty or incorrect instructions. It is for these reasons that all methods introduced into the validation program are now taken through the sequence given above.

DESIGN OF COLLABORATIVE TESTS

Traditionally, analytical methodology has been validated through the process of collaborative testing. The collaborative test is designed so that qualified participants (collaborators) each make one or more measurements on identical samples using the same test method. Then, from a statistical analysis of the results, an estimate is made of the within-laboratory and between-laboratory precision of the test method. If the samples are of know concentration (unknown, of course to the analyst), or if a material can be supplied for analysis that simulates a known sample concentration, then an estimate of the accuracy of the method can be obtained. This general technique has been used very widely for the validation of methods for the analysis of water, drugs, food and agricultural products, fertilizers, coal, ores, etc.

When collaboratively testing a stationary source emission method, the sampling procedure, as well as the analytical aspects of the method, must be evaluated. This usually means that the participants must sample a real source representative of those where the method will be used. However,

depending upon the physical state of the material being sampled, this may create a series of complex problems with no easy solutions. For the test to be successful all participants must have access to the same pollutant concentration in the stack, for, if they cannot obtain identical samples, they surely will not get reproducible results. For gaseous pollutants, this can frequently be accomplished by extracting a side stream from the stack and piping it to ground level, where it is delivered through a manifold. The collaborators simultaneously sample the gaseous pollutant through ports on the manifold. Although a manifold could be constructed to accommodate a relatively large number of collaborators, participation is usually limited to ten or less because of coordination problems, the great expense of maintaining personnel in the field, and the relative scarcity of qualified collaborators for any particular test.

Attempts to collaboratively test methods for pollutants that exist in particulate form become complicated by the requirement that all test teams sample the material isokinetically directly from the stack. Here the problem becomes one of the simultaneous extraction of representative samples from the stack by each of the collaborative test teams. Since spatial and temporal variations may constantly be occurring in both the velocity profile and the pollutant profile, an attempt must be made to compensate for this so that each participant has access to statistically identical or equivalent samples. In the first particulate tests, this was attempted by allowing each test team to sample at each traverse point in the stack for the same period of time during the 2-hour run, although in any given time period each team would be sampling at a different point.[11,12,13] For circular stacks, this automatically limited participation to four test teams, each sampling independently through one of the

90-degree ports, and each rotating to the next port on a signal from the test coordinator. It was reasoned that, over the entire sampling period, each collected sample should be representative of the average stack pollutant concentration for that time period. However, since the estimates of between-laboratory variability are based upon the differences observed among collaborators within each sampling run, these estimates would be affected to the extent that the samples are nonrepresentative in character.

Due to the nature of the sampling procedures and the requirements for simultaneous sampling by all collaborators, 2 weeks was the minimum time in which a collaborative test of some source emission methods could be conducted. It is difficult to find a test site where unit operation is essentially constant for that length of time, due to load demand changes or the possibility of process upsets. Such changes can affect the pollutant loading from run to run during the test. Thus, the collection of true replicate samples on consecutive runs becomes almost impossible, and a more indirect approach was used to estimate the variability within laboratories on repeated measurements. This was done by grouping the determinations (runs) into blocks of approximately equal average concentration using the most appropriate blocking criteria available. (Such blocking criteria are based upon unit operating parameters, which would be expected to influence the emission levels, such as fuel feed rates, power generation levels, production rate, raw material feed rates, and electrostatic precipitator voltages. Also, opacity data from in-stack transmissometers have been used along with unit operating parameters in test data blocking.) The within-laboratory standard deviation estimates were then calculated based upon the variability of each collaborator within these blocks. But, while this procedure did tend

to reduce the influence of process changes on these precision estimates, it could not eliminate their effects completely, and in some cases this was reflected in within-laboratory estimates that are somewhat higher than would otherwise be obtained.

Because of problems cited in the last two paragraphs, a new and improved approach was sought to the collaborative testing of methods for pollutants that exist in particulate form. The objective was to develop a test design that would allow sampling by a greater number of collaborators and that would not be affected by the random variations in the velocity and pollutant concentration profiles mentioned above. The result was a new test design using paired sampling trains in which two probe-pitot tube assemblies could simultaneously sample at very nearly the same point in the stack. Since the paired probe tips sample in rather close proximity, this greatly minimizes the effects of spatial and temporal stack variation on the samples collected by the adjacent probes. In addition, this allows the extraction of up to eight individual samples per run on a circular stack, with a resultant increase in the number of degrees of freedom for the statistical analysis.

The test design specified a 3-week sampling period with six independent test teams operating separate trains in three of the paired train systems for the entire duration of the test. Both trains in the remaining pair were operated by a single team, with one operator running both meter boxes. Since all equipment in each train in this pair was virtually identical, had been carefully calibrated, and was operated by the same individual, then the sample pair collected during any given run could be considered replicates. The participation of the collaborator that operated this pair of trains was restricted to 1 week, with a

different team participating in this capacity during each of the 3 weeks. Thus, four pairs of samples were obtained on each run -- one sample pair by a single collaborating laboratory, and three sample pairs by three pairs of laboratories. At the end of each 30-minute sampling interval, each paired train assembly and test team rotated to an adjacent port in the stack so that, at the conclusion of each run, each team and train had sampled an equal time at each traverse point.

Estimates of the variability within a laboratory were based upon the differences in concentration reported by the paired-train laboratory for the replicate samples on each run. Differences among laboratories were estimated by contrasts between paired trains that were operated by the six single-train laboratories. This test design has been applied to a collaborative test of EPA Method 5 to be discussed in an ensuing section.[14]

Collaborative testing of source emission methods suffers from two restrictions, that are not found in testing methods for other materials. Most important of these is the limited number of participants. Attempts to compensate for this by taking a greater number of samples only partially solves the problem; i.e., with only four laboratories participating, an equipment malfunction or a deficiency in the performance of just one laboratory can have a very adverse effect on the outcome of the test. This obstacle has been largely overcome by the paired-train test design discussed above, but at a great increase in cost. A less serious restriction concerns the limited pollutant concentration range that can be examined in a collaborative test at a real source. Of course, with cooperation from plant personnel, some range in concentration can be obtained by varying conditions such as precipitator voltages, excess air, etc., and the test can sometimes be augmented by standard cylinder

gases covering a range of concentration. But only rarely can the complete applicable range of a method be investigated using real samples at a single test site. Also, the use of such standard gases is usually the only means by which method accuracy can be evaluated, since the true concentrations of pollutants in the stack are rarely, if ever, known. However, since one cannot duplicate in a cylinder gas the environmental conditions and possible interferences that could exist in some stack gas streams, such cylinder gas data must be regarded as the best accuracy of which the method is capable under ideal conditions.

ANALYSIS OF COLLABORATIVE TEST DATA

Before discussing the results obtained from the collaborative tests, a brief discussion of the information available from collaborative testing, and the manner in which this information is derived is in order. A primary purpose of the test is the determination of the precision components of the method, i.e., how closely a user can expect to repeat his results on subsequent application of the method on identical samples and how closely different users can expect to agree when analyzing separate but identical samples. These precision components are estimated using either a coefficient of variation approach or an analysis of variance technique after first performing suitable data transformations when necessary.

Prior to evaluating the precision of the method, the determinations are tested for equality of variance using Bartlett's test[15] for horogeniety of variances. In addition, the determinations are passed through two common variance stabilizing transformations, the logarithmic and the square root, and Bartlett's test is again applied. The use of transformations

serves two purposes. First, it can put the data into an acceptable form for an analysis of variance; and second, it can provide information concerning the true nature of the distribution of the sample points. The transformation that provides the highest degree of run equality of variance is accepted and used in deriving the precision estimates.

Acceptance of the logarithmic transformation implies that there is a proportional relationship between the true mean, δ, and the true standard deviation, σ, and that the ratio of the standard deviation to the mean (the coefficient of variation, β) remains constant.

$$\beta = \frac{\sigma}{\delta}$$

Once this relationship has been established, the data may be analyzed in its linear form and the standard deviations presented as a coefficient of variation times an unknown mean, δ, i.e.

$$\sigma = \beta\, \delta.$$

Alternately, an analysis of variance may be performed on the transformed data, and the components of variance then converted back to the linear form to provide uniform coefficient of variation estimates for the determinations.

When the distributional nature of the data is such that its original or linear form provides the highest degree of equality of variance, then this implies that there is a constant variance that is independent of the mean level. In this case, the variances are estimated by a pooled analysis of variance on the original data.

In order to provide the maximum useful information, the test must be designed and the data analyzed in such a fashion that the precision estimates for a determination can be partitioned into its respective

variance components. The variance components of interest are those that estimate the variability within a laboratory, the overall variability between laboratories, and that portion of the overall variability that is due to the individual biases of different laboratories.

The within-laboratory standard deviation, σ, measures the dispersion in replicate single determinations made by one laboratory team (same field operators, laboratory analysts, and equipment) sampling the same concentration level. Simply stated, this is the measure of a laboratory's ability to repeat its own test results when all experimental factors and relevant environmental conditions are held constant. This term has also been referred to as the standard deviation of repeatabilty or, more simply, "repeatability," and carrys with it the concept of making repeated measurements on the same sample, or on identical samples. This value is estimated from within each collaborator-block combination or from replicate samples collected by the same laboratory using paired sampling trains.

The between-laboratory standard deviation, σ_b, measures the total variability in a determination due to simultaneous determinations by different laboratories sampling the same true stack concentration, μ. The between-laboratory variance, σ^2_b, may be expressed as

$$\sigma^2_b = \sigma^2_L + \sigma^2$$

and consists of a within-laboratory variance plus a laboratory bias variance,σ^2_L. The between-laboratory standard deviation is estimated using the run results or the within-run differences between paired sampling trains operated by different laboratories. This term estimates the degree of agreement to be expected among different laboratories who have independently collected and analyzed identical samples. The between-laboratory

standard deviation is frequently called standard deviation of reproducibility, or "reproducibility".

The laboratory bias standard deviation, $\sigma_L = \sqrt{\sigma^2_b - \sigma^2}$, is that portion of the total variability that can be ascribed to differences in the field operators, analysts, and instrumentation, and to different manners of performance of procedural details left unspecified in the method. This term measures that part of the total variability in a determination that results from the use of the method by different laboratories, as well as from modifications in usage by a single laboratory over a period of time. The laboratory bias standard deviation is estimated from the within-laboratory and between-laboratory estimates previously obtained.

Before leaving this section, it is appropriate to say something about how method accuracy is expressed. With respect to the accuracy of a method we attempt to define its absolute accuracy; i.e., how well does the measurement value agree with the actual or true value. As stated previously, estimates of method accuracy must frequently be based on the analysis of standard cylinder gases. One approach is to have each collaborator measure the concentration of the cylinder gas (or other material), after which a mean and a standard deviation are calculated for the group of collaborators. A 95 percent confidence interval is then calculated around this mean. If the true concentration of the cylinder gas lies within this 95 percent confidence interval, then the method is said to be unbiased and accurate within the limits of its precision. A more common means of stating method accuracy consists of averaging the respective biases of all collaborators and expressing this as a percentage (either positive or negative) of the overall mean, or of the true value, when known. Both approaches to stating method accuracy will be found in the various collaborative test reports.

SECTION 4

RESULTS OF THE METHODS STANDARDIZATION PROGRAM

Since the initiation of the program in August 1972, evaluations and collaborative studies have been conducted on a number of methods. While the overall aim of the project is the standardization of these methods, the evaluations, collaborative tests, and subsequent data analysis have been structured to determine both the strong and weak points of the methods. By determining those areas of weakness in a given method, recommendations have been made for changes that will improve the accuracy and precision of that method. The actual collaborative testing phase of the program began with a test of EPA Method 7 for oxides of nitrogen (NO_x) in December 1972[16] and has more recently included tests of Method 9 (opacity) in October 1974[17] and Method 5 (particulates) in September 1975.[14] Table 1 lists those methods for which some collaborative testing has already been completed, and a discussion of the results of these investigations will now follow.

STACK GAS VELOCITY AND VOLUMETRIC FLOW RATE

Collaborative tests of the Type S Pitot Tube Method (EPA Method 2 - Ref. 1, pp. 24884-24885) were conducted in conjunction with tests of EPA Method 5 at three sites: a Portland cement plant, a coal-fired power plant, and a municipal incinerator.[18] There were 15, 16 and 12 traverses at the three respective sites and four collaborating laboratories at each. The data from one laboratory at the power plant site were not used, and some determinations were not made due to equipment failure during the sampling runs. This resulted in a total of 150 separate determinations of both velocity and volumetric flow rate for use in the data analysis.

The runs at each site were grouped into blocks based upon the velocity heads. The precision components were shown to be proportional to the mean of the determinations and are expressed as percentages of the true mean as shown in Table 2 for both the velocity and the volumetric flow rate determinations.

A more recent test of Method 2 was conducted at a different municipal incinerator, and included 13 runs by six different collaborators.[14] Test design and data analysis were similar to those used for the above studies, as were the resulting precision estimates.

Based upon the results of these tests, the precision of the volumetric flow rate determination seems adequate for use with other test methods in determining pollutant emission rates. The small σ_L indicates that the method is inherently rugged; i.e., it is not subject to large biases from one user to another. A previous single-laboratory study indicated that for nonturbulent streams, Method 2 provides an accurate estimate of the true stack gas velocity at the higher velocities of 55 to 60 feet per second.[19] Relative accuracy is somewhat less at velocities of about 30 feet per second.

GAS ANALYSIS FOR CARBON DIOXIDE, EXCESS AIR, AND DRY MOLECULAR WEIGHT

Collaborative tests of the Orsat methodology for the determination of CO_2, excess air, and stack gas molecular weight[20] were conducted in conjunction with the three tests of EPA Method 5 to be discussed in a subsequent section. The Orsat procedure tested was similar to that of EPA Method 3 (Ref. 1, pp. 24886-24887) with one important exception.

Method 3 required that the analysis of a gas sample be repeated until three consecutive analyses that vary no more than 0.2 percent by volume for each component being analyzed are obtained. In these tests, the average of three consecutive analyses was used, but the requirement that they differ by no more than 0.2 percent by volume was not enforced. This was a very significant deviation from Method 3, and the test schedules were such that the results may be questioned. (See subsequent discussion of Method 5 tests.) The results will therefore not be reported here.

Five other collaborative tests have been conducted to investigate various aspects of the Orsat methodology.[21] Four of these were field studies in which four to seven collaborators analyzed replicate samples from a larger bulk sample of combustion effluent gas. The number of replicate analyses allowed varied according to the design of the experiment, and ranged from four to seven. Under these restrictions, none of the collaborators met the Method 3 operator performance criterion of three consecutive analyses that differ by no more than 0.2 percent by volume for each component, so the results are not relatable directly to Method 3.

The fifth test in this series was a laboratory study in which seven collaborators analyzed replicate samples from an EPA stationary source simulator facility. Three different levels of carbon dioxide and oxygen were studied, and only those values that met the performance criterion for a valid Method 3 analysis were used in the data analysis. From these results, between-laboratory standard deviation for Method 3 in the range of 0.20 to 0.39 percent CO_2, and 0.38 to 0.55 percent O_2 were obtained for these two components, depending on the level tested. Within-laboratory

standard deviations were not calculated because repeated measurements of sets of three analyses that met the Method 3 performance criterion were not made.

The most recent collaborative field test of Method 3 was conducted at a municipal incinerator in conjunction with a test of Method 5, and consisted of 13 runs with up to seven collaborators sampling per run.[14] A revised version of Method 3 was used for this test. The operator performance criterion of the revised method states that the analysis must be repeated until the molecular weight for three consecutive analyses differs from their mean by no more than 0.3 gram/gram-mole.

Precision estimates were obtained for the various parameters using an ANOVA approach and are summarized in terms of standard deviation in Table 3. In addition, the Orsat CO_2 data were examined to estimate the magnitude of the error that might be introduced when a determined particulate concentration is corrected to 12 percent CO_2. The between-laboratory standard deviation was 0.40 percent CO_2 by volume at the levels encountered at this test site. If the true CO_2 level were 2.3 percent, then two independent laboratories might be expected to obtain values of 2.1 and 2.5 percent, respectively. For two laboratories that had determined the same particulate concentration, this would result in a 19 percent difference in the reported particulate concentration after correction to 12 percent CO_2.

Based upon the results of all studies completed, it is concluded that: (1) the Orsat method is tedious and requires great attention to detail and technique; (2) the original EPA Method 3 operator performance criterion was not easily met in the field, and even meeting this criterion does not ensure that highly reproducible and accurate results will be

obtained; (3) the use of Orsat data to routinely convert particulate catches to such reference conditions as 12 percent CO_2 and 50 percent excess air can introduce significant errors into the corrected particulate loading; and (4) the Orsat is quite satisfactory for use in determining stack gas molecular weights.

MOISTURE FRACTION WITH USE OF METHOD 5

Collaborative tests of the procedure for determination of moisture fraction (in conjunction with EPA Method 5 - Ref. 1, pp. 24888-24890) have been conducted at a Portland cement plant, a coal-fired power plant, and a municipal incinerator, using four sampling teams carrying out 15, 16, and 12 sampling runs, respectively, at the three sites.[20] The absence of several values from the data set necessitated using runs as repetitions, and undoubtedly caused the error term to be inflated due to run-to-run variation in stack moisture content. Other factors, as discussed in the succeeding section on particulates likely adversely affected the results. An analysis of variance procedure on this data produced an estimated within-laboratory standard deviation of 0.032, a between-laboratory standard deviation of 0.045, and a laboratory bias standard deviation of 0.032.

A more recent test of a revised version of Method 5 was conducted at a second municipal incinerator. This test consisted of 13 runs over a 3 week period with eight trains sampling per run.[14] The data were submitted to statistical analysis using an ANOVA model. A two-way model without interaction was used to avoid blocking the runs, and the run by train interaction was used for the error term. This test design and data analysis resulted in estimates of 0.009, 0.012, and 0.008 for the moisture

fraction within-laboratory, between-laboratory, and laboratory bias standard deviations, respectively. These are considerably better than those previously reported and are probably more representative of the true performance capabilities of the method.

PARTICULATES

Collaborative tests of EPA Method 5 (Ref. 1, pp. 24888-24890) for determination of particulate matter emissions were conducted at a coal-fired power plant,[11] a Portland cement plant,[12] and a municipal incinerator.[13] Four sampling teams participated in each test, accomplishing 16, 15, and 11 runs, respectively, at the three sites. At the cement plant and the incinerator, sampling was performed through four ports located at 90-degree angles on the circular stacks. The power plant sampling was done through four ports in a horizontal duct leading directly to the stack. In an attempt to ensure collection of statistically equivalent and representative samples by all participants, each team sampled at each traverse point in the stack for the same period of time during the course of each 2-hour run. This required that the teams sample simultaneously, each sampling through a different port and then rotating to an adjacent port at each quarterly time interval until all four ports had been sampled by each team.

For the purpose of statistical treatment, the determinations were grouped into blocks using the most appropriate blocking criteria that could be devised for each test. A coefficient of variation approach was then used to calculate a within-laboratory, between-laboratory, and laboratory bias component for each test. These ranged from 25.3 to 31.1 percent, 36.7 to 58.4 percent, and 19.6 to 51.0 percent, respectively, for the three tests.

It immediately becomes obvious that these estimates indicate Method 5 to have relatively poor precision. However, it had been shown by other single-laboratory studies that under very carefully controlled conditions, using multiple-probes sampling simultaneously over a very small area, with well designed equipment, highly competent personnel who execute all operations in an identical and representative manner, etc., that Method 5 is capable or giving precise and reliable results.[22] So before accepting these results as being representative of the true capabilities of Method 5, a few factors were examined that could have contributed to this apparent imprecision.

The first factor concerns the limited number of participants that could be accomodated, as has been mentioned earlier. While we attempted to find fully qualified people to participate in these tests, four collaborative teams per test represents a very small statistical population upon which to base our conclusions. Thus, any bias or deficiency in the performance of a single team has a very significant effect on the apparent precision of the method.

Another factor that might have contributed to the apparent imprecision of Method 5 is that of collaborator fatigue. Because of the very considerable expense involved in running these tests, it was decideded that two sampling runs per day would be made in order to collect, within a 2-week period, the 12 to 16 samples per collaborator required for a meaningful statistical treatment of the data. With the uprigging and downrigging of the equipment of four test teams, the movement of this equipment around the stack during port changes, performance of Orsat analyses, etc., excessively long days of 12 to 14 hours were often required. While such work days may not be uncommon in source testing, it is abnormal to maintain such schedules for the duration of time required for a collaborative test. It is thus possible

that participant fatigue may have had some adverse effects on method precision.

The tests were designed so that during each run the collaborators rotated from port to port, each sampling the same points in the stack over the course of each 2-hour run, though not at the same point at the same time. With this pattern, it was hoped that any effects due to spatial and temporal variations in the stack particulate concentration would be randomized out and that all participants would statistically be able to collect identical and representative samples. However, we doubt that we completely eliminated all effects due to spatial and temporal concentration variations, and these could be reflected to some extent in the precision.

At the time of these tests, it was difficult to find sites with the necessary facilities and with personnel who would voluntarily cooperate in having their plants used for such a program. Therefore, the selected test sites were frequently less than desirable from the standpoint of port location, distance of sampling point from flow disturbances, velocity profile, control equipment, pollutant concentration range, etc. Also, it could not be required that the plant maintain steady-state conditions over the duration of the collaborative test as could be required in compliance testing. For example, the sampling at the power plant was conducted in a horizontal duct under conditions that did not really conform to EPA Method 1. And the particulate loading at the cement plant varied by a factor of eight over the 2-week period of the test. Obviously, this might make the simultaneous collection of representative samples by the various test teams more difficult.

At the time of designing and performing these tests, it was believed that Method 5, as promulgated, was written in sufficient detail to assure that different users would execute it in a proper and reproducible manner. Therefore, participants in these tests were allowed to use Method 5 in

accordance with their exact, individual interpretations of the method's instructions without outside influence from the test coordinator. However, it is now apparent that the method lacked sufficient clarity in some critical areas, and some test teams lacked the experience necessary for its proper application. So, it may be possible that the inclusion of more detail into the method would have improved its precision.

Because of problems and uncertainties in the original test designs, a fourth collaborative test of Method 5 was undertaken using the paired sampling train test design previously discussed.[14] The test, conducted at a municipal incinerator in September 1975, used a revised and more detailed version of Method 5 since the original method write-up was considered deficient. At the same time, the philosophy of conducting collaborative testing was changed to some extent. First, potential collaborators were screened more carefully to ensure that only well experienced and competent personnel would be selected to participate in the test. And the role of the test supervisor was increased in order to assure that the collaborators operated within the constraints of the revised Method 5 and the associated Methods 2 and 3. Thirteen runs were accomplished over a 3-week period, with one run per day, eight samples per run.

The data analysis for the within-laboratory precision estimate was based upon the differences in concentration reported by the paired-train laboratory for the replicate samples on a given run. The standard deviation estimated from the pooled data of all three laboratories is 13.8 mg/m^3, for a coefficient of variation of 10.4 percent of the average determined concentration (Table 2). The laboratory bias standard deviation was estimated by ANOVA from the contrasts between paired trains operated by the six single-train laboratories. This estimate was 8.15 mg/m^3, which

gives a coefficient of variation of 6.1 percent. Combining these gives a between-laboratory standard deviation estimate of 16.0 mg/m^3, or 12.1 percent of the mean. There was no detectable effect among these results due to spatial and/or temporal changes in the stack flow.

These test results show the precision capabilities of Method 5 to be considerably greater than had been previously thought, and this may be due in part to the improved test design. Spatial and temporal source effects were eliminated; the testing followed a more relaxed pace; and the larger population of participants was a definite advantage. In addition, much tighter control was exercised over the actions of the collaborators than was previously done. The participating test teams were probably among the best in this country; nevertheless, three of the meter boxes were found to be outside the allowable specifications for dry gas meter calibration according to the revised method when checked at the test site. Had not the test plan provided for calibration checks on-site (and recalibration if necessary) the outcome of the test would surely have been adversely affected. Other problems were observed during the test itself, and these were called to the attention of the collaborators for correction. Such corrective actions had not been taken in the former tests.

It is impossible to determine the exact effect on the test of using the revised method. However, since the revision does supply much of the needed detail in Method 5, it is safe to assume that the effect was positive. In fact, if anything is to be learned from these studies, it is that successful execution of Method 5 requires care and close attention to detail. In the hands of a competent test team who will use such care and attention, Method 5 is capable of giving satisfactory and precise results, as shown by this most recent study.

SULFUR DIOXIDE

EPA Method 6 for SO_2 (Ref. 1, pp. 24890-24891) was evaluated,[19] and then was collaboratively tested at two different sites, a 140-megawatt coal-fired electric generating plant and an oil-fired pilot combustion plant.[16] A randomized block design was employed at each site, with four different blocks of emission concentration levels that ranged from about 232 mg/m^3 to about 1750 mg/m^3. These blocks, each of which consisted of four runs sampled at 60-minute intervals, were obtained on consecutive days. The intent was to maintain a constant true SO_2 emission concentration level at the sampling points on the four runs within each block to permit an accurate determination of the within-laboratory precision of Method 6. Samples at the power plant were collected from a manifold through which a stream of the stack gas was delivered to rooftop level, and SO_2 concentration levels were varied by the injection of dilution air upstream of the sampling ports. At the pilot plant, samples were collected directly from the duct downstream of the furnace and heat exchanger, and concentration levels were varied by doping the system with SO_2. Each run involved the simultaneous collection of an exhaust sample over about a 20-minute period by each of four collaborating laboratories through their assigned ports.

In addition to the above experiments, two auxillary tests were also conducted at both sites to complement the real-sample data obtained. The first of these was a gas cylinder accuracy test to provide an independent assessment of the accuracy of Method 6. This test involved three different standard cylinder gases containing mixtures of SO_2 in nitrogen, the concentrations of which had previously been determined by the supplier with a claimed accuracy of ± 1 percent. On each of the test days, each

collaborator obtained one sample from each cylinder according to the Method 6 procedure to be later analyzed with the day's collaborative test samples. The second auxillary test involved the triplicate analytical determination of the SO_2 concentrations implicit in four unknown standard sulfate solutions to isolate the accuracy and precision of the sample analysis phase of Method 6.

An analysis of the collaborative test data using a coefficient of variation approach provided estimates of the precision components listed in Table 2. From these values, it is evident that Method 6 is capable of good precision when used by competent personnel. Analysis of the standard sulfate solution produced standard deviations of 1.1, 2.4, and 2.2 percent of the mean value, respectively, for the within-laboratory, between-laboratory, and laboratory bias terms. In comparing these values with the tabled values for the entire method, it becomes obvious that most of the precision variation resides in the field sampling phase of Method 6, as opposed to the analytical phase.

The gas cylinder accuracy test showed Method 6 to be accurate at SO_2 concentrations up to about 480 mg/m^3, but indicated that it acquires a significant negative bias above the range of about 480 to 800 mg/m^3. The apparent bias was found to be in the field sampling phase rather than in the analytical phase of the method. This conclusion was based on the fact the collaborators reported values for the high-level cylinder gases that were generally lower than those claimed by the gas supplier. However, it is now thought that this conclusion of negative bias was incorrect, and that the low reported values probably resulted from decay of the cylinder SO_2 concentration or some other unknown phenomenon. More recent work within EPA[23] has indicated that collection efficiency is

practically 100 percent and the method is unbiased up to SO_2 concentrations of at least 5000 mg/m^3.

NITROGEN OXIDES

EPA Method 7 for NOx (Ref. 1, pp. 24891-24893) was evaluated for interference effects in the laboratory[19] and then subjected to collaborative testing at the same two sites used for the Method 6 tests described above.[16] A third test was conducted at a nitric acid plant which utilizes a proprietary catalytic ammonia oxidation process.[24] The tests were based on a randomized block design similar to that described above for SO_2. Tested concentrations ranged from about 160 mg/m^3 to about 2400 mg/m^3, expressed as NO_2. Auxillary tests at the three sites included the sampling of standard cylinder gases at the coal-fired power plant and the pilot combusion plant, and the sampling of a standard test atmosphere at the nitric acid plant set up and controlled by personnel of the National Bureau of Standards. Four collaborators participated in each test. In addition, the collaborators were given a series of unknown potassium nitrate standard solutions to be analyzed with the samples.

The data from the first two tests were pooled to provide a larger data base and then analyzed using a coefficient of variation approach. A similar analysis was performed on the nitric acid plant data. The resulting precision estimates are presented in Table 2, first for the pooled power plant/pilot combustion plant data, and then for the nitric acid plant data. Note that the estimates for the nitric acid plant study were uniformly higher, roughly by a factor of two. Because of the larger data base resulting from the pooling of the data from the first two tests, and because of the frequently unstable conditions encountered at the

nitric acid plant, it is felt that more reliance may be placed on the precision estimates obtained from the former tests.

An analysis of variance on the nitrate solution data disclosed that nearly all of the analytical laboratory-to-laboratory variance component is attributable to day-to-day variation in laboratory measurements instead of to significant laboratory biases. Analysis of the standard test atmosphere established that Method 7 is unbiased and accurate within the limits of its precision.

SULFURIC ACID MIST/SULFUR DIOXIDE

EPA Method 8 (Ref. 1, 24893-24895) for the measurement of sulfuric acid (H_2SO_4) mist (including any free SO_3) and sulfur dioxide (SO_2) was collaboratively tested at a dual absorption contact process sulfuric acid plant with a rated capacity of 900 tons of acid per day.[25] Simultaneous samples were collected by four collaborative test teams (in a manner analogous to that previously described for two of the Method 5 tests) through four ports located at 90-degree angles in the stack.

The collaborative test plan called for the collaborators to obtain 16 samples during a 2-week period. The sampling was curtailed by inclement weather, and as a result only 14 sampling runs were made. The collaborators were also provided with standard sulfuric acid solutions to be analyzed along with their field test samples as described in previous tests.

An inspection of the collaborative test data revealed that H_2SO_4 mist concentrations in this test varied by as much as an order of magnitude between collaborators within single runs, with several high values that were of a magnitude to suggest that they were not representative of the true concentration in the stack. Sulfur dioxide determinations showed a

similar variation, varying by as much as a factor of two. A correlation analysis of the test data showed a significant negative correlation between the H_2SO_4 mist determinations and the SO_2 determinations, i.e., high values reported for acid mist were associated with low SO_2 values at a greater frequency than could be expected by chance alone.

The data from the test were arranged in blocks and analyzed both with and without the inclusion of six extraordinarily high acid mist values and their corresponding SO_2 values. The precision components shown in Table 2 for H_2SO_4 mist and in Table 3 for SO_2, were developed after these values were excluded from the data set. It is immediately obvious that the precision of the acid mist determination was extremely poor in this test even after the elimination of the six most extreme values. Considering that the tested SO_2 concentrations ranged from about 480 mg/m^3 to about 800 mg/m^3, it is also apparent that the SO_2 determination, while better than the acid mist, was not highly precise.

The results from the analyses of the unknown sulfate solutions were used to evaluate the accuracy and precision of the analytical phase of the method separate from the field sampling phase. For the analytical phase of the method, the within-laboratory standard deviation was found to be independent of the mean level, and was estimated as 3.51 mg SO_2/m^3. The between-laboratory standard deviation and the laboratory bias standard deviation were determined to be proportional to the mean level, and were estimated as 3.7 percent of δ and 3.5 percent of δ, respectively. The analytical phase was shown to be accurate, within the precision of the method, at all three levels of concentration studied. These levels ranged from 254 to 1,073 mg/m^3 equivalent SO_2 concentration.

From the precision estimates given above, it is quite evident that the predominant sources of error were in the field sampling phase of the

test. Because of the significant negative correlation between the H_2SO_4 mist determinations and the SO_2 determinations, one is immediately led to suspect some intrinsic problem in the method such as an inability to satisfactorily separate the SO_2 and the acid mist fractions of the sample. But, with limited data from only one test, it is impossible to say at present whether the imprecision observed is due to a real deficiency in the method, some unknown phenomenon peculiar to the test site, or to other factors such as those discussed in the preceding section on Method 5. We are presently planning additional work to determine the true reliability of Method 8.

OPACITY OF STACK EMISSIONS

Collaborative testing of EPA Method 9 (Ref. 1, p. 24895) for visual determination of opacity of emissions from stationary sources was conducted using certified observers, to obtain data that would allow statistical evaluation of the method.[17] Three collaborative test sites were used: a training smoke generator, a sulfuric acid plant, and a fossil fuel-fired steam generator. The initial test on the training smoke generator was conducted to provide background information on the use of the method, while the test at the sulfuric acid plant and the fossil fuel-fired steam generator were conducted to obtain information on the use of the method on applicable sources under field conditions. At no time during any of the tests were warm-up or practice runs allowed prior to the test itself.

These tests required the determination of average opacity, defined as the average of 25 readings taken at 15 second intervals. For the purpose of this study, one set of 25 readings was designated a "run". The collaborators began taking readings on a signal from the test supervisor, and thereafter at 15

second intervals until the required 25 observations were obtained. Concurrent with the observers' readings, plume opacity readings were taken from the in-stack transmissometer. The accuracy of the method was judged by the deviations of the observers' readings from the actual opacity as measured by this calibrated in-stack transmissometer.

Five separate tests of EPA Method 9 were conducted, if both the white smoke and the black smoke phases of the training generator study are considered as comprising one test. For each test, Table 4 presents the pertinent information on the number of runs completed, the number of observers participating, and the range of opacity studied. The studies were deliberately restricted to the lower opacity ranges within which the EPA standards lie.

While the smoke generator and the sulfuric acid plant tests were designed to evaluate the accuracy and precision of Method 9 as written, the steam station studies, in addition to this, were designed to investigate the effects of various factors on the performance of the method. The experimental factors studied included the angle of observation and the relative experience of the observer. Variations to the method to be evaluated included reading in 1 percent rather than in 5 percents, and using the average responses of two observers as opposed to a single observer's result to determine whether these yielded increased accuracy. The observers at each test were divided into two groups for the test, a control and an experimental. The control group observed the plume at all times from a position consistent with the method as written and read in increments of 5 percent. The experimental group either read the plume from a more extreme angle in increments of 5 percent or from the same angle as the control but in increments of 1 percent. Each group was composed both of observers who had considerable field experience with the method and

of observers who had relatively little such experience.

Due to the adverse sky and wind conditions during Tests 1 and 2 at the steam station, not all of the planned evaluations were useful. There was an inability to read the low opacity plumes against the type of background that existed, and as a result, the determinations were generally well below the concurrent meter average. The precision estimated, however, is independent of the accuracy of the determination. Separate precision estimates were therefore developed for these tests, and for the tests at the other sites. Composite estimates based upon the results of all the tests were also derived, and because the individual estimates were similar from one test to another, only the composite estimates shown in Table 3 will be presented in this report. Using data from the training generator and from Test 3 at the steam station, a composite estimate of the accuracy of Method 9 was derived for ideal (clear-sky) conditions. This estimate compares the expected deviation of the observer from the average metered opacity and is given by the equation, deviation = 3.13 - 0.31 (meter average), for the range from 5 to 35 percent average opacity. As the equation indicates, observers tend to read slightly high at the very low opacities, exhibit good accuracy at around 10 to 15 percent average opacity, and acquire a definite negative bias at the higher opacities.

With respect to the other experimental factors and variables studied, it was concluded from the clear-sky data of Test 3 that (1) the angle of observation does affect the observer's determinations, and in this study, the most accurate readings were made when the group was at an approximately 45 degree angle to the sun; (2) the experienced observers were able to read average opacity more accurately than the inexperienced observers, but the difference occurred mainly in the higher opacity range (≥25 percent); (3) the 1 percent increment data exhibited greater within-observer variability and was less accurate than the

5 percent increment data; and (4) averaging the results of two observers yielded increased accuracy over the result of a single observer. Based partly on the results of these studies, Method 9 was revised and improved and has now been repromulgated to replace the original method of 1971.[26]

CARBON MONOXIDE

A collaborative test of EPA Method 10 for carbon monoxide (CO) (Ref. 2, pp. 9319-9321) was carried out at a petroleum refinery, where seven collaborators sampled the emissions from the CO boiler downstream of the fluid catalytic cracking unit catalyst regenerator.[27] All collaborators simultaneously sampled through a manifold connected to the CO boiler stack using the integrated sampling option of Method 10. Each collaborator obtained four 60-minute samples per day until 16 runs were completed at each of two CO concentration levels. In addition to the stack samples, each collaborator analyzed six standard cylinder gases (CO in nitrogen) that had been supplied for the test by the National Bureau of Standards.

It was the intent of the experimental design to maintain the CO concentration constant for the 16 runs at each of the two concentration levels (blocks) so that readings within each block could be considered replicates for the purpose of calculating the within-laboratory precision component. However, it was found that the blocks could not be physically maintained at constant concentration, so an indirect approach based upon the pairing of runs of similar concentration was used to estimate the within-laboratory standard deviation of Method 10. This value, and the estimated value for the between-laboratory and laboratory bias standard deviations are given in Table 3. From an analysis of the data for the NBS standards, a somewhat similar between-laboratory term was calculated (26 mg/m^3 as compared to the 32 mg/m^3 shown in Table 3 for the field data). However, the standards data showed about a threefold

improvement in the within-laboratory standard deviation over the field data (5.2 mg/m^3 vs 14 mg/m^3), and this is probably due to the presence of some source variability in the field estimates.

In analysis of the NBS standards, collaborators differed in the amount of bias exhibited, and the average bias was dependent on the CO levels. In general, a sizeable positive bias was shown at the lower CO levels, but a negative bias was evident at the highest CO level. Method 10 as executed in this study produced results with only moderate accuracy of $\pm$ 101 mg/m^3 (2σ level) on the average over the concentration range of 277 to 1048 mg CO/m^3. One factor that adversely affected the accuracy of Method 10 is that most commercial NDIR instruments have a significant amount of curvature in the calibration curves, and many of the collaborators did not adequately correct for this nonlinearity of response. Another factor is the calibration gases themselves, since some calibration gas suppliers provided certificates of analysis that showed errors of as much as 30 percent when compared with the NBS standard gases.

BERYLLIUM

The EPA beryllium method (Method 104 - Ref. 3, pp. 8846-8850) was collaboratively tested in a process plant where different beryllium ceramic products are manufactured -- a process that involves machining, grinding, blending, priming, forming, and polishing.[28] Air from the process is continuously exhausted through a series of HEPA filters before entering the 3-by-5-foot stack from which sampling was done simultaneously by four collaborators. This collaborative test comprised 13 runs, each on a different day, where four different collaborative organizations sampled simultaneously over the same 30-point traverse, with each point being sampled 8 minutes by each collaborator. The emission levels of beryllium in the stack sampled were low, being in the

neighborhood of one-tenth that of the permissible standard emission rate.

Three types of samples were prepared by the National Bureau of Standards specifically for this collaborative test: filters with beryllium oxide, ampoules with suspended beryllium oxide, and ampoules with soluble beryllium in 0.25 molar hydrochloric acid. These samples were given to the collaborators at the field site to be later analyzed with the field samples in their home laboratories.

There were three statistical analyses performed. The primary one was a two-way analysis of variance to obtain the variance of repeated observations per collaborator and to obtain the variance between collaborators. A secondary analysis was the same except beryllium-loading results were used in place of the emission rate results. The third analysis, was to determine if the average velocity per sampling point per run correctly represented the geometrical variance in velocity throughout the test run even though they were measured at different times.

The precision estimates for the emission rate data are given in Table 2. Estimates derived from the beryllium-loading results were virtually identical to these, so it is evident that the velocity and volumetric flow rate measurements did not contribute significantly to the imprecision observed in the emission rate data. It appears that almost all the differences between collaborators during a run were due to differences in the solution (wash plus impinger contents) portion of the samples. Three of the four collaborators did not differ significantly in the amount of beryllium collected per run on their filters. Since, on the average, about 77 percent of the beryllium was collected from the solution portion of the sample (probably from the nozzle and probe washes), it is likely that the sample clean-up was a major source of error. This would be compounded by the fact that beryllium concentrations were extremely low at this test site.

The collaborators relative precision in the measurement of beryllium from the NBS standard samples was considerably greater than for their field samples

(within-laboratory standard deviation of about 10 percent), but the standard samples contained larger amounts of beryllium. However, analysis of these standard samples indicated a definite collaborator bias, which in general was proportional to the beryllium level, and, on the average, was about 20 percent negative. The average bias on the filter samples was essentially zero, but only because large negative and positive biases cancelled out. One collaborator exhibited essentially no bias on any of the sample types, and one laboratory measured the filter concentrations without bias. Since one collaborator always managed to measure beryllium without bias and since bias was sometimes positive and sometimes negative, it is apparent that the observed bias is a property of the collaborators rather than being inherent in the method itself. Thus, because of questionable competency of some of the collaborators, it is unlikely that the true performance capabilities of Method 104 were determined by this test.

TABLE 1. METHODS COLLABORATIVELY TESTED UNDER THE METHODS STANDARDIZATION PROGRAM

Parameter	Method of determination	EPA Method No.[a]
Stack gas velocity and volumetric flow rate	S-type pitot tube	2
Stack gas molecular weight and CO excess air	Orsat	3
Stack gas moisture content	Condensation and volumetric measurement	5
Particulates	Dry filtration and gravimetric determination	5
Sulfur dioxide	Selective absorption and barium thorin titration	6
Nitrogen oxides	Phenol disulfonic acid	7
Sulfuric acid mist/sulfur dioxide	Selective absorption and barium thorin titration	8
Opacity of stack	Visual estimation of percent opacity	9
Carbon monoxide	Nondispersive infrared absorption	10
Beryllium	Filtration/impingement and atomic absorption	104

[a] Methods 2, 3, 5, 6, 7, 8, and 9 are described in Reference 1, Method 10 in Reference 2, and Method 104 in Reference 3.

TABLE 2. PRECISION ESTIMATES FOR THOSE PARAMETERS WHERE STANDARD DEVIATION WAS PROPORTIONAL TO THE MEAN VALUE, δ

Method No.	Parameter, units	Standard deviations, percent of mean value (δ)[a]		
		σ	σ_b	σ_L
2	Velocity, ft/sec	3.9	5.0	3.2
2	Volumetric flow rate, ft^3/hr	5.5	5.6	1.1
5	Particulate matter, mg/m^3	10.4	12.1	6.1
6	SO_2, mg/m^3	4.0	5.8	4.2
7[b]	NO_x, mg/m^3	6.6	9.5	6.9
7[c]	NO_x, mg/m^3	14.9	18.5	10.5
8	H_2SO_4 mist (including SO_3), mg/m^3	58.5	66.1	30.8
104	Be, g/day	43.5	57.7	37.9

[a] σ = within-laboratory deviation; σ_b = between-laboratory deviation, σ_L = laboratory bias.

[b] Pooled power plant/pilot combustion plant data.

[c] Nitric acid plant data.

TABLE 3. PRECISION ESTIMATES FOR THOSE PARAMETERS WHERE STANDARD DEVIATION WAS INDEPENDENT OF THE MEAN VALUE.

Method No.	Parameter, units	Standard deviation parameter units[a]		
		σ	σ_b	σ_L
3	CO_2, percent	0.20	0.40	0.35
3	O_2, percent	0.32	0.61	0.52
3	Dry mol. wt, g/g-mole	0.035	0.048	0.033
5	Moisture fraction	0.009	0.012	0.008
8	SO_2, mg/m^3	123	115	99
9	Opacity, percent	2.05	2.42	1.29
10	CO, mg/m^3	14.3	32.3	29.0

[a] σ = within-laboratory deviation; σ_b = between-laboratory deviation; σ_L = laboratory bias.

TABLE 4. COLLABORATIVE TEST OF EPA METHOD 9

Site/test	Number of runs	Number of observers	Opacity range, percent
Training generator:			
White smoke	20	9	0-35
Black smoke	16	9	0-35
Sulfuric acid plant	30	11	0-15
Steam station/test 1	10	10	0-30
Steam station/test 2	18	10	0-25
Steam station/test 3	24	8	0-40

SECTION 5

REFERENCES

1. U.S. Environmental Protection Agency. Standards of Performance for New Stationary Sources. Federal Register. 36(247):24876-24895. 1971.

2. U.S. Environmental Protection Agency. Standards of Performance for New Stationary Sources. Federal Register. 39(47):9308-9323, 1974.

3. U.S. Environmental Protection Agency. National Emission Standards for Hazardous Air Pollutants. Federal Register. 38(66):8820-8850, 1973.

4. Mitchell, W. J. and M. R. Midgett. Improved Procedure for Determining Mercury Emissions from Mercury Cell Chlor-Alkali Plants. Jour. APCA. (In Press.) U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

5. Knoll, J. E. and M. R. Midgett. Determination of Hydrogen Sulfide in Refinery Fuel Gases. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. (In Press.)

6. U.S. Environmental Protection Agency. Standards of Performance for New Stationary Sources. Federal Register. 40(152):33157-33166, 1975.

7. U.S. Environmental Protection Agency. National Emission Standards for Hazardous Air Pollutants. Federal Register. 40(248):59549-59550, 1975.

8. Mitchell, W. J. and M. R. Midgett. Adequacy of Sampling Trains and Analytical Procedures Used for Fluoride. Atmos. Environ. (In Press.) U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

9. Youden, W. J. Statistical Techniques for Collaborative Tests. The Association of Official Analytical Chemists, Washington, D. C., 1973. pp. 33-36.

10. Mitchell, W. J. and M. R. Midgett. Means to Evaluate Performance of Stationary Source Test Methods. Environ. Sci. Technol. 10(1):85-88, 1976.

11. Hamil, H. F. and R. E. Thomas. Collaborative Study of Method for the Determination of Particulate Matter Emissions from Stationary Sources (Fossil Fuel-Fired Steam Generators). U.S. Environmental Protection Agency Research Triangle Park, North Carolina. Report No. EPA-650/4-74-021. 1974.

12. Hamil, H. F. and D. E. Camann. Collaborative Study of Method for the Determination of Particulate Matter Emissions from Stationary Sources (Portland Cement Plants). U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-74-029. 1974.

13. Hamil, H. F. and R. E. Thomas. Collaborative Study of Method for the Determination of Particulate Matter Emissions from Stationary Sources (Municipal Incinerators). U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-74-022. 1974.

14. Hamil, H. F. and R. E. Thomas. Collaborative Study of Particulate Emissions Measurements by EPA Methods 2, 3, and 5 Using Paired Particulate Sampling Trains (Municipal Incinerators). U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-600/4-76-014. 1976.

15. Dixon, W. J. and F. J. Massey, Jr. Introduction to Statistical Analysis, 3rd Ed. New York. McGraw-Hill. 1969.

16. Hamil, H. F., D. E. Camann, and R. E. Thomas. The Collaborative Study of EPA Methods 5, 6, and 7 in Fossil Fuel-Fired Steam Generators, Final Report. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-74-013. 1974.

17. Hamil, H. F., R. E. Thomas, and N. F. Swynnerton. Evaluation and Collaborative Study of Method for Visual Determination of Opacity of Emissions from Stationary Sources. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-75-009. 1975.

18. Hamil, H. F. and R. E. Thomas. Collaborative Study of Method for Determination of Stack Gas Velocity and Volumetric Flow Rate in Conjunction with EPA Method 5. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-74-033. 1974.

19. Hamil, H. F. Laboratory and Field Evaluations of EPA Methods 2, 6, and 7. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-74-039. 1973.

20. Hamil, H. F. and R. E. Thomas. Collaborative Study of Method for Stack Gas Analysis and Determination of Moisture Fraction with Use of Method 5. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-74-026. 1974.

21. Mitchell, W. J. and M. R. Midgett. Field Reliability of the Orsat Analyzer. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Jour. APCA. 26(5):491-495. 1976.

22. Mitchell, W. J. and M. R. Midgett. Method for Obtaining Replicate Particulate Samples from Stationary Sources. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-75-025. 1975.

23. Knoll, J. E. and M. R. Midgett. The Application of EPA Method 6 to High Sulfur Dioxide Conentrations. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. (In Press.)

24. Hamil, H. F. and R. E. Thomas. Collaborative Study of Method for the Determination of Nitrogen Oxide Emissions from Stationary Sources (Nitric Acid Plants). U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-74-028. 1974.

25. Hamil, H. F., D. E. Camann, and R. E. Thomas. Collaborative Study of Method for the Determination of Sulfuric Acid Mist and Sulfur Dioxide Emissions from Stationary Sources. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-75-003. 1974.

26. U.S. Environmental Protection Agency. Standards of Performance for New Stationary Sources. Federal Register. 39(219):39874-39876. 1974.

27. Constant, P. C. Jr., G. Sheil, and M. C. Sharp. Collaborative Study of Method 10 - Reference Method for the Determination of Carbon Monoxide Emissions from Stationary Sources - Report of Testing. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-75-001. 1975.

28. Constant, P. C. Jr., and M. C. Sharp. Collaborative Study of Method 104 - Reference Method for Determination of Beryllium Emission from Stationary Sources. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Report No. EPA-650/4-74-023. 1974.